

Use Case 21: Detecting shifts in cell type composition in complex tissues using lineage specific markers and reference epigenomes from the Epigenome Atlas

American Society of Human Genetics
Boston, MA

October 22, 2013

Presented by the
Bioinformatics Research Laboratory

Baylor
College of
Medicine

Summary of Use Case 21

Background: Methylation profiling reflects the average level of cytosine methylation of the sampled cell population in different genomic regions. Complex tissue samples are typically composed of several different cell types, and different samples from the same tissue may have different proportions of constituent cell types. These differences often lead to variability in estimating methylation levels in different samples from same the tissue. In breast cancer, for example, immune cell infiltration is often observed. True differential methylation changes between cases and controls may thus be masked by differences that result from different proportions of constituent cell types found in breast cancer. Comparison of experimental breast cancer samples with reference epigenomes from the Human Epigenome Atlas provides an approach to help identify shifts in cell type composition in complex tissues.

Results: Comparing 450k array breast tumors and normal breast samples with reference methylomes in the Epigenome Atlas using heatmap tool, we find that 450k array tumor samples are more epigenetically similar to immune cell samples from the Atlas than the normal breast 450k samples. This indicates possible increase in proportion of immune cells in the tumor samples, which is consistent with previous studies.

Use case dataset

- Dedeurwaerder, S. et al. (2011) "Evaluation of the Infinium Methylation 450K technology", Epigenomics 3(6):771-84.
 - 16 breast tissue samples
 - Profiled using Illumina 450k array
 - 8 normal breast samples
 - 8 primary breast tumor samples
- NIH Roadmap Epigenome Data: <http://www.genboree.org/epigenomeatlas/multiGridViewerPublic.rhtml>
 - 45 MeDIP-seq (16 different cell types)
 - 108 RRBS (67 different cell types)
 - 57 WGBS (45 different cell types)
 - **Please see slide 42 for introduction to MeDIP, RRBS, WGBS, and 450k array**

Genomic regions used in comparisons

- **Background:** Some CpGs that are interrogated by the Illumina 450k array are methylated/unmethylated specifically in one cell lineage in the Epigenome Atlas. These lineage specific CpGs are much more informative about cell type composition of a complex tissue sample, than the other CpGs. In order to reduce the noise introduced by having uninformative CpGs in the comparison, we will select and use in the comparison only the lineage specific CpGs in the 450k array.
- **Results:** Cell lineages were identified by hierarchical clustering of all cell types in the Atlas using different epigenomic marks (H3K4me1, H3K4me3, etc), followed by selection of clusters of cell types that were consistently observed. 18 cell lineages were selected this way. We then quantified the level of methylation in each methylome track for the regions probed by the 450k array. A comparison between level of methylation in each cell lineage with cells not in that lineage was then performed for each of the regions probed in the 450k array using the LIMMA tool. CpGs that were consistently methylated/unmethylated in cells from a lineage (i.e. myeloid), yet exhibited the opposite methylation status in all other lineages, were considered lineage specific CpGs, and were used to generate the ROI track we will use in this experiment.

Marker CpGs resource

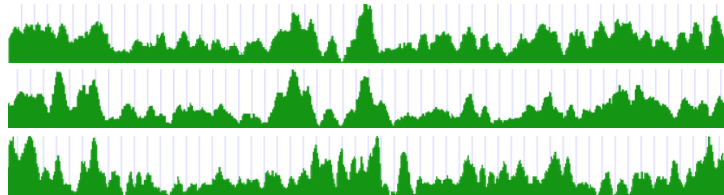
- On Genboree there are public available sets of CpGs that we selected as markers for cell lineages.
 - 450k array CpGs
 - 27k array CpGs
 - Overlapping 450k and 27k CpGs
- LIMMA tool on Genboree can be used to select your own set of lineage specific loci

Methodology Overview

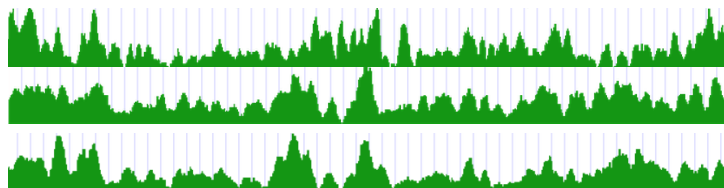
Select Atlas data tracks: Use “View Track Grid” tool to select methylome tracks from the Epigenome Atlas generated by each experiment type

Select 450k array data tracks: Use “View Track Grid” tool to select methylome tracks from the Epigenome Atlas generated by each experiment type

Epigenome Atlas
Methylomes

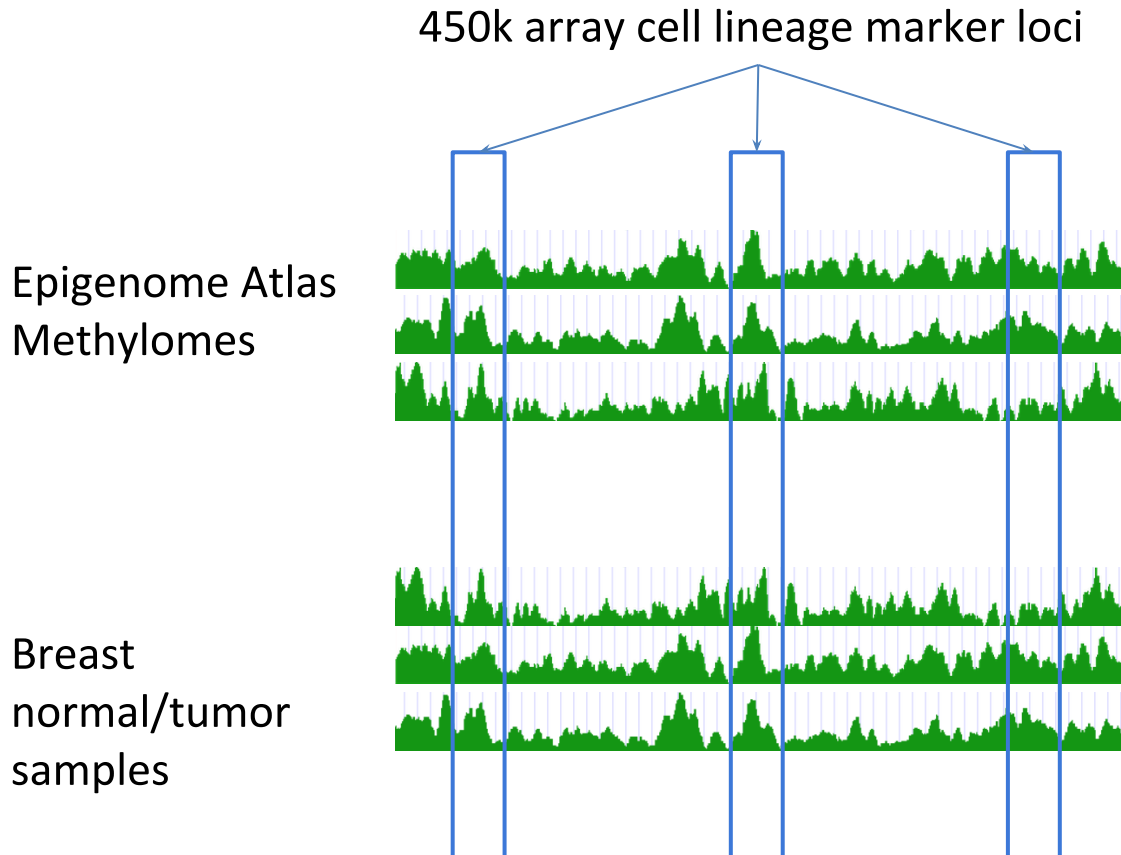


Breast
normal/tumor
samples



Methodology Overview

Extract methylation levels: Calculate the average level of methylation in each of the chosen loci for each input track. This is done automatically by the heatmap tool.



	Locus 1	Locus 2	Locus 3
Cell type 1	0.8	0.7	0.3
Cell type 2	0.7	0.6	0.7
Cell type 3	0.6	0.8	0.2
	Locus 1	Locus 2	Locus 3
	1		
Breast 1	0.6	0.8	0.2
Breast 2	0.8	0.7	0.3
Breast 3	0.7	0.8	0.2

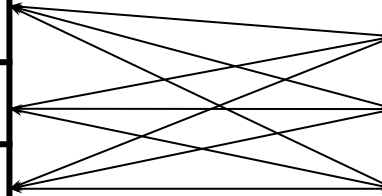
Methodology Overview

Epigenome Atlas Methylomes

	Locus 1	Locus 2	Locus 3
Cell type 1	0.8	0.7	0.3
Cell type 2	0.7	0.6	0.7
Cell type 3	0.6	0.8	0.2

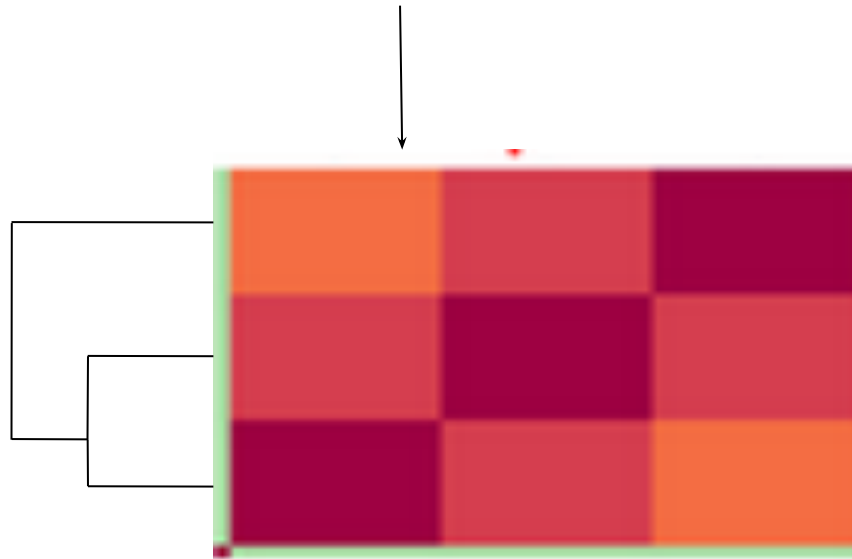
Breast normal/tumor samples

	Locus 1	Locus 2	Locus 3
Breast 1	0.6	0.8	0.2
Breast 2	0.8	0.7	0.3
Breast 3	0.7	0.8	0.2

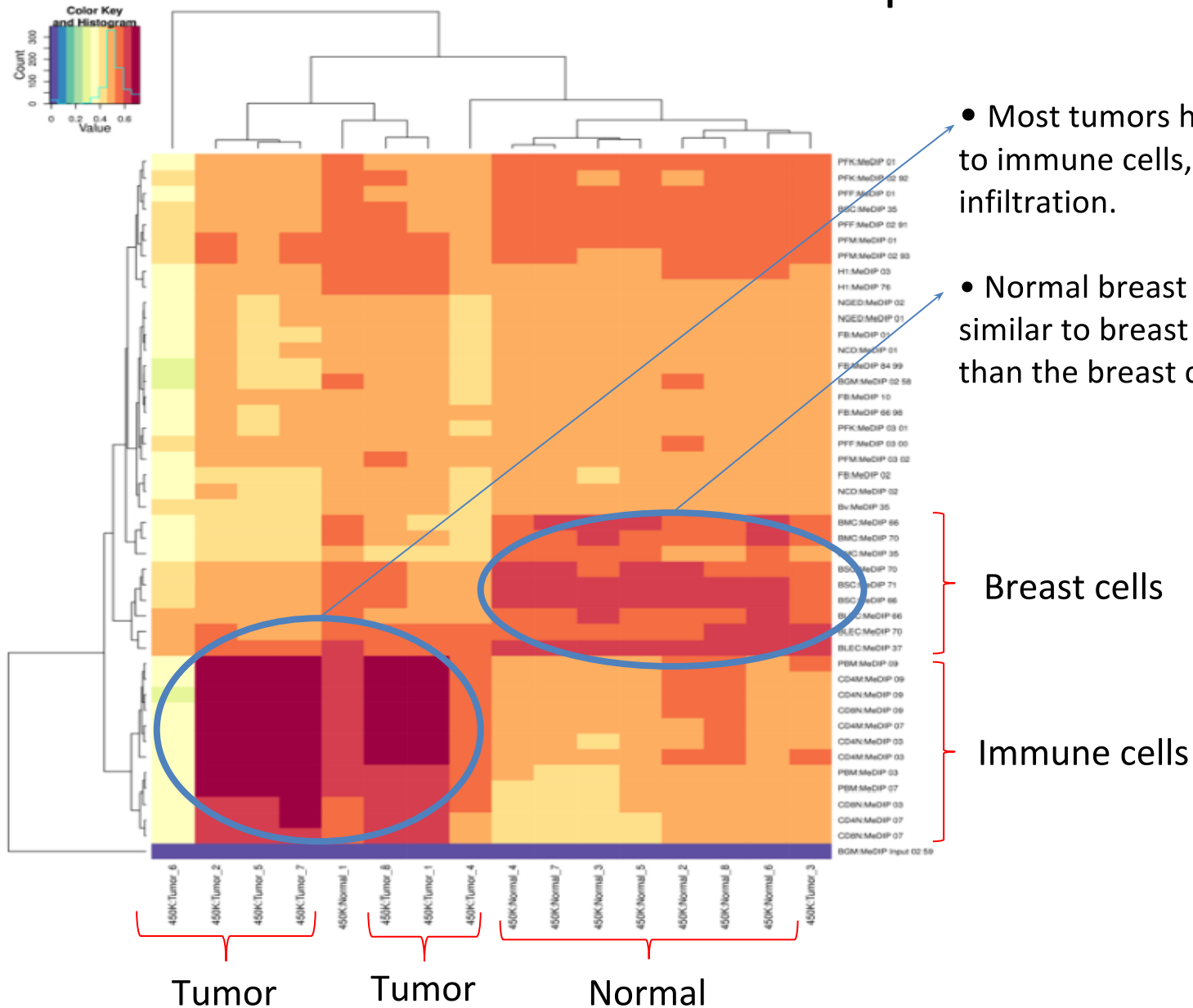


Compute similarity between samples: Calculate Euclidian distance between all pairs of samples, being one breast 450k sample and one methylome from the Atlas. Also performed automatically by heatmap tool.

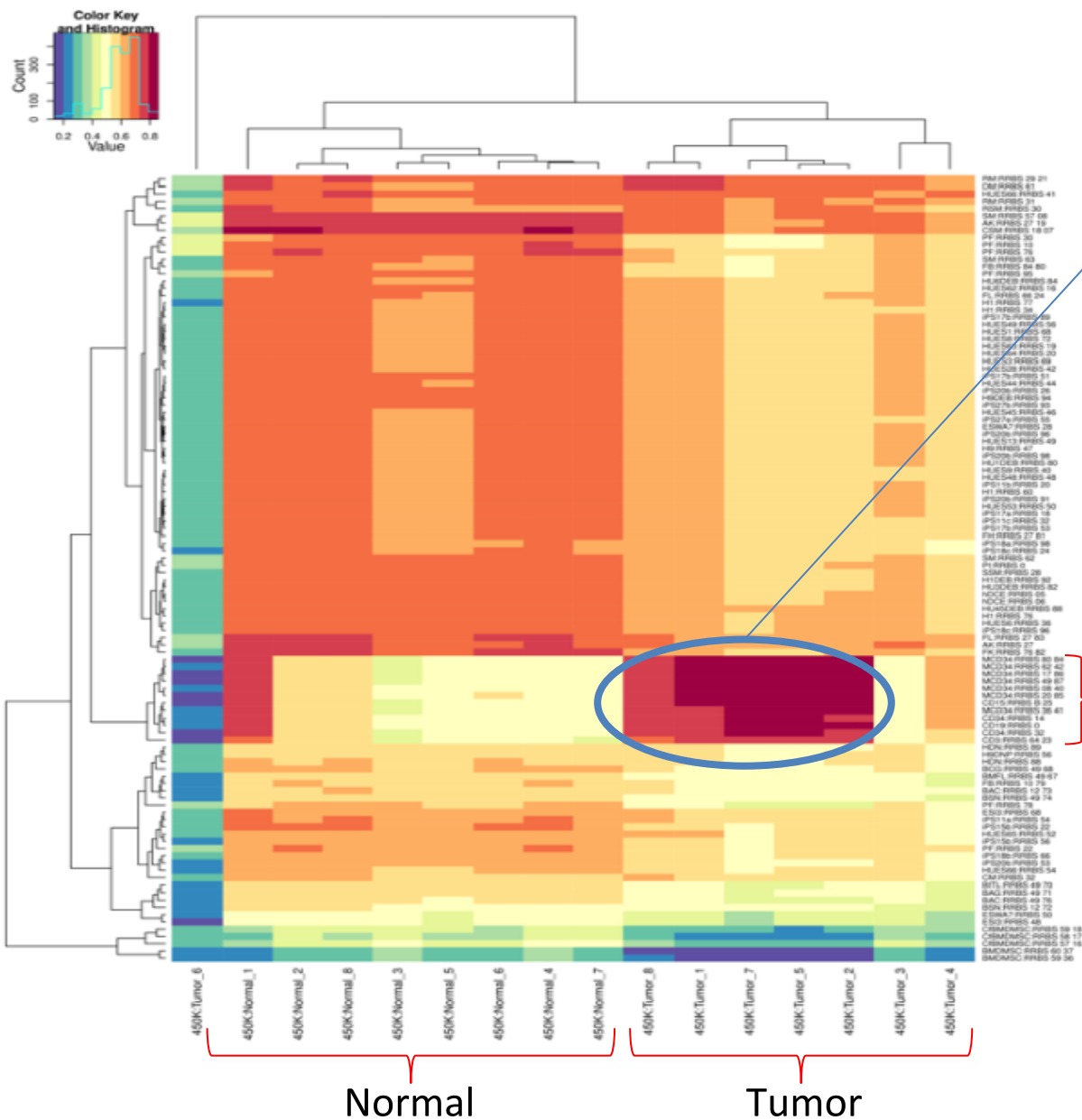
Hierarchical clustering
and heatmap plotting



Expected results: Comparison Breast450k vs Atlas MeDIP-seq

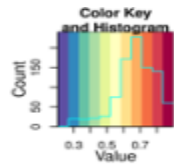


Expected results: Comparison Breast450k vs Atlas RRBS



- Possible immune cell infiltration in the tumors can also be observed in the comparison with RRBS atlas samples

Immune cells



New Genboree Users - Slides 13-17 provide steps for new Genboree users on how to create a database and a project page.

Existing Genboree Users - If you have attended past Genboree Workshops or are familiar with the Genboree Workbench then you may briefly review these slides and start on slide 18 for the actual use case

The Genboree Workbench: Web-based Data Management & Analysis

The screenshot shows the Genboree Workbench interface. At the top, there is a navigation bar with tabs: System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. The 'Data' tab is selected. Below the navigation bar, a welcome message reads 'Welcome to the Genboree Workbench! [Getting Started]'. The main area is divided into three sections: 'Data Selector', 'Details', and 'Input Data' and 'Output Targets'.

Data Selector: Various Data Types (tracks, files, and ROIs (region of interests), etc)

Details: Specific information on files/samples selected in the "Data Selector"

Input Data: Tells the tool to use this input data/file

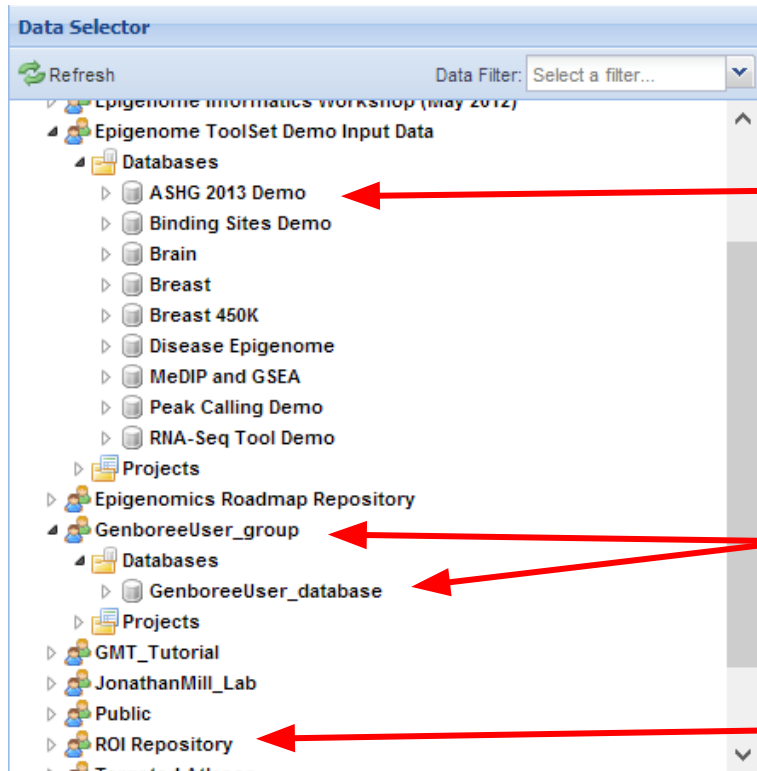
Output Targets: Tells the tool where to deposit results

The 'Data Selector' section shows a tree view of data sources, including 'genboree.org' and 'genboree.bcgsc.ca'. The 'Details' section shows a table with columns 'Attribute' and 'Value', containing information like 'Name' and 'Description'. The 'Input Data' and 'Output Targets' sections show icons for adding, removing, and saving data.

Important: Toolset Menu turns **GREEN** when "Input Data" and "Output Targets" are properly populated for a tool to run. Please note that "System/Network" and "Help" options are always green since "User Profile", "Jobs", and "Request Feature" are always available for use and do not need "Input Data" and "Output Targets" to be populated.

Preparation Prior to Starting the Use Case

- “**GenboreeUser_group**” is a name template for an automatically created Genboree user group **for you** where “**GenboreeUser**” is **your user name**.
- Similarly, “**GenboreeUser_database**” is a name template for your database.
- Of course, you may create many more databases and may create and be member of many other groups.



Under “Epigenome Toolset Demo Input Data” you will find “ASHG 2013 Demo” database, where we have provided you with sample data to try out the use cases

When making screenshots for providing instructions we have used “GenboreeUser_group” and “GenboreeUser_database” as output targets, however, you will use your own group and database. Following slides will show you how to create database and project

ROI Repository contains database of annotated regions (eg. Gencode and Refseq annotations)

Displaying Tool Setting help options in the Workbench



System/Network Data QC and Pre-processing Genome Transcriptome Cistrome Epigenome Metagenome Visualization Help

Welcome to

Data Selector

- Databases
- Entity Lists
- Entrypoints

- Create Database
- Rename Database
- Delete Database

Details

Attribute	Value
-----------	-------

Input Data

Help: Tool Settings

Help: Create Database

Configure Tool

This tool is a recent addition. Please contact genboree_admin@genboree.org with questions or comments, or for help using it on your own data.

This tool will create/add a new database in the target Group. Note that the database to be created should NOT already exist in the group.

Output Targets

Instructions:

- Drag 1 destination group into "Output Targets". The new database will be created in this group.

Output type(s):

- Group
min: 1 ; max: 1

Can be empty? NO

Tool-Specific Settings

Settings:

- Reference Sequence**
Select the genome assembly the database will use.
- Database Name**
The name of the database to be created/added. [REQUIRED]
- Description**
A line or two describing the database. [OPTIONAL]
- Species**
Automatically selected based on Reference Sequence. [OPTIONAL]
- Version**
Automatically selected based on Reference Sequence. [OPTIONAL]
- Submit**
Once you've reviewed the name of the new database to be created.

Bioinformatics Research Laboratory
Baylor College of Medicine.

is available **free for academic use**.

HGSC
HUMAN GENOME SEQUENCING CENTER

A grey background (not green) means that the tool is not active. Clicking a non-active tool displays the help text that includes instruction for how to activate the tool.

To create a database, you need to drag something in "Output Targets". Instruction says to drag a **Group** into "Output Targets"

Steps for Creating a Database

Step I - Drag **your** group from "Data Selector" into "Output Targets". GenboreeUser_group and GenboreeUser_database is placeholder for your group and databases

Step II - Click "Data" => "Databases" => "Create Database". "Create Database" tool **IS** active, since it requires a "Group" to be in "Output Targets". Select "Create Database" for tool settings.

Step III - Select "Template: Human (hg19)"

Step IV - Type database name (i.e. "GenboreeUser_database" and click "Submit"

Tool Settings: Create Database

Target Group:

Reference Sequence: Template: Human (hg19)
Template: Human 3/12 (Hg15)
Template: Human 3/12 (Hg16)
Template: Human chr 12(Hg15)
Template: Human chr 12(Hg16)

Database Name:
Description:
Species:
Version:

Submit **Cancel**

Output Targets

GenboreeUser_group



Steps for Creating a Project page

GENBOREE **BCM**

System/Network Data QC and Pre-processing Genome Transcript

Welcome to bench! [Getting Started]

Data Selector

Refresh

genboree.org

- Atlas Tool
- BRL AUTO
- EDACC
- Epigenome Informatics Demo Output Data
- Epigenome Informatics Workshop (May 2012)
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- GenboreeUser_group**
- GMT_Tutorial
- JonathanMill_Lab
- Public
- ROI Repository
- Targeted Atlases

Databases

- Entity Lists
- Entrypoints
- Files
- Projects**
- Samples & Sample Sets
- Tracks

Create Project

Select a filter...

Create Project

Tool Overview

Target Group:

Group: GenboreeUser_group

Settings

Project Name **Unique Name**

Project Title ASHG workshop

Project Description Result output from ASHG Wks

Submit Cancel

Output Targets

GenboreeUser_group

Step I - Drag your group from "Data Selector" into "Output Targets".

Step II - "Create Project" tool **IS** active, since it requires a "Group" to be in "Output Targets". Select "Create Project" for tool settings.

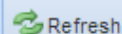
Step III - Type Project Name, Title, and Desc. and click "Submit".

NOTE: Project name has to be unique for all Genboree users, so you could do something like 'Use_case_21' + your initials (i.e. "Use_case_21-abc")

Step IV - Select "Refresh" in "Data Selector" to view your newly created database and project page in your group.

Welcome to the Genboree Workbench! [Getting Started]

Data Selector



Data Filter: Select a filter...

genboree.org

- Atlas Tools Access
- BRL AUTO TEST
- EDACC
- Epigenome Inform
- Epigenome Inform
- Epigenome Tools
- Epigenomics Roadmap Repository
 - Databases
 - Data Freeze 1 - Full Repo
 - Data Freeze 2 Repository
 - Release 5 Repository
 - Release 6 Repository
 - Release 7 Repository
 - Release 8 Repository
 - Release 9 Repository
 - Projects
 - GenboreeUser_group
 - GMT_Tutorial
 - JonathanMill_Lab
 - Public
 - ROI Repository

Step 1 - Drag "Release 9 Repository" database from "Epigenomics Roadmap Repository" to "Input Data"

Step 2 - Select "View Track Grid"

- View Track Grid
- View Sample Grid
- Tabular Annotation Viewer
- Launch UCSC Genome Browser

Attribute	Value
Group	Epigenomics Roadmap Repository
Role	public
Name	Release 9 Repository
Description	Release 9 Repository
Species	Homo sapiens

Input Data



Release 9 Repository

Output Targets



Select how you want the tracks displayed in the “View Track Grid” tool.

The screenshot shows the 'View Track Grid' tool settings window. It has a title bar 'Tool Settings' and a close button. The main title is 'View Track Grid' with a help icon. There is a '+ Tool Overview' section. Below it is a 'Databases with tracks of interest:' section containing a 'Database:' field with the value 'Release 9 Repository' and a 'Group:' field with the value 'Epigenomics Roadmap Repository'. Below that is a 'Settings' section. In the 'Settings' section, the 'X-axis attribute' is set to 'eaAssayType' and the 'Y-axis attribute' is set to 'eaSampleType'. Both dropdown menus are highlighted with red boxes. Red arrows point from these boxes to two separate red boxes on the right. The top red box contains the text 'Step 3 - Select “eaAssay Type”' and the bottom red box contains the text 'Step 4 - Select “eaSample Type”'. Below the 'Settings' section is an 'Advanced Settings' section. At the bottom of the window are 'Submit' and 'Cancel' buttons, both highlighted with red boxes.

Tool Settings

View Track Grid

+ Tool Overview

Databases with tracks of interest:

Database: Release 9 Repository Group: Epigenomics Roadmap Repository

Settings

X-axis attribute: eaAssayType

Y-axis attribute: eaSampleType

Page Title: Grid Viewer: Tracks from Relea

Grid Title: Tracks from Release 9 Reposit

X Label: eaAssayType

Y Label: eaSampleType

+ Advanced Settings:

Submit Cancel

Step 3 - Select “eaAssay Type”

Step 4 - Select “eaSample Type”

Releases

Information

Save Track Selections

Forums

Contributors

Choose a group and database to save selections in:

Select a Group:

This is the group where your selections will be saved

GenboreeUser_group

Select a Database:

Choose a database within your group to save to

GenboreeUser_database

Save Selection as:

Enter a name to identify this set of selections

UC_Breast450k_WGBS_Atlas

Save Selections

Cancel

Step 7: Select your user group ("GenboreeUser" is used here for illustration purposes)

Step 8: Select the database that you created earlier

Step 9: Name this list of tracks

Step 10: Click on Save Selections

- [Data Access Policy](#)
- Data embargo period: from 04/15/2013 - 01/15/2014 or earlier as specified [here](#)
- Select cells by **clicking and dragging**, then use "View Selections" in the Selections menu to
- Use "Save Selections" in the Selections menu to save selected (highlighted) cells in a group
- To see data authors, other metadata, and to download data, click a sample name in the first
- Expression Array data may be downloaded [here](#)
- Human Epigenome Atlas releases are intended to be cumulative: e.g. Release 3 includes all
- **NOTE:** Some pages may not be accessible over low bandwidth internet connections. This page

Tracks from Release 9 Repository

Filter rows: Selections

eaAssayType

eaSampleType

	Bisulfite-Seq	MeDIP-Seq	MRE-Seq	RRBS	5	5	5	5	5	5	1
Adipose Derived Mesenchymal Stem Cells					3	1					1
Adipose Nuclei					2	2	4	2	2	1	2
Adipose Tissue											
Adrenal Gland											
Adult Kidney				2		2		2	2	2	2
Adult Liver	1				2	4	4	4	4	5	2
Aorta	1				2	2	2	2	2	2	1

The screenshot shows the Genomex application interface. On the left, a table displays genomic data for various cell types. The columns are labeled 'eaSampleType', 'eaAssayType', and several assay types: 'Bisulfite-Seq', 'MedIP-Seq', 'MRE-Seq', and 'RRBS'. A red arrow points to the 'MedIP-Seq' column header. The table rows include 'Breast Fibroblast Primary Cells', 'Breast Luminal Epithelial Cells', 'Breast Myoepithelial Cells', 'Breast Stem Cells', 'Breast vHMEC', 'CD14 Primary Cells', and 'CD15 Primary Cells'. A 'Save Selections' dialog box is open in the center, prompting the user to choose a database (GenboreeUser_database) and save the selection with a name (UC_Breast450k_MeDIP_Atlas). The dialog box also has 'Save Selections' and 'Cancel' buttons.

eaSampleType	Bisulfite-Seq	MedIP-Seq	MRE-Seq	RRBS
Breast Fibroblast Primary Cells				
Breast Luminal Epithelial Cells	1	3	5	
Breast Myoepithelial Cells	1	3	3	
Breast Stem Cells		4	4	
Breast vHMEC		1	1	
CD14 Primary Cells				1
CD15 Primary Cells				1

Welcome to the Genboree Workbench! [\[Getting Started\]](#)

Data Selector

Refresh

- genboree.org
 - Atlas Tools Access
 - BRL AUTO TEST
 - EDACC
 - Epigenome Informatics Demo Output Data
 - Epigenome Informatics Workshop (May 2012)
 - Epigenome ToolSet Demo Input Data
 - Databases
 - Binding Sites Demo
 - Brain
 - Breast
 - Breast 450K**
 - Disease Epigenome
 - MeDIP and GSEA
 - Peak Calling Demo
 - RNA-Seq Tool Demo
 - Projects
 - Epigenomics Roadmap Repository
 - GenboreeUser_group
 - GMT_Tutorial
 - JonathanMill_Lab
 - Public

Step 13: Drag "Breast 450K" database to "Input Data"

Details

Attribute	
Group	Epigenome ToolSet Demo Input Data
Role	subscriber
Name	Breast 450K
Description	Template for Human Genome, UCSC Build Hg19

Input Data

-
- Breast 450K

Step 14: Click on "View Track Grid"

- View Track Grid
- View Sample Grid
- Tabular Annotation Viewer
- Launch UCSC Genome Browser

Step 16: Select Selections > Save selections

- Select cells by clicking and dragging, then use the "View Selections in" pulldown in the top
- **NOTE:** Some pages may not be accessible over low bandwidth internet connections. This page

Tracks from Breast 450K

Filter rows: **Selections** ▾ Choose Databases

eaAssayType" →

450K

eaSampleType

Normal	8
Tumor	8

**Step 15: Select all
16 tracks: 8 Normal
and 8 Tumor**

Save Track Selections

**Choose a group and database to save
selections in:**

Select a Group:

This is the group where your selections will be saved

GenboreeUser_group ▾

Select a Database:

Choose a database within your group to save to

GenboreeUser_database ▾

Save Selection as:

Enter a name to identify this set of selections

UC_Breast450k

Save Selections

Cancel

**Step 17: Select your user group
("GenboreeUser" is used here for
illustration purposes)**

**Step 18: Select the database
that you created earlier**

**Step 19: Name
this list of tracks**

**Step 20: Select "Save
Selections"**

Welcome to the Genboree Workbench! [Getting Started]

Note: Your saved track lists will be in the chosen database under Lists & Selections/Lists of Tracks

- GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections**
 - Lists of Tracks**
 - UC13_450K_Normal
 - UC13_450K_Tumor
 - UC13_Breast_450K_1
 - UC13_Breast_450K_2
 - UC13_Breast_MeDIP_Atlas
 - UC_Breast450k
 - UC_Breast450k_MeDIP_Atlas**
 - UC_Breast450k_RRBS_Atlas
 - UC_Breast450k_WGBS_Atlas
 - UseCase12_Breast_RRBS
 - UseCase12_Breast_RRBS_Sun_2011
 - UseCase12_Epi_Atlas_MeDIP_Seq
 - UseCase13_450K_Normal
 - UseCase13_450K_Tumor
 - UseCase1 Brain_A

Details

Attribute	Value
Track 1	BGM:MeDIP 02 58
Track 2	BGM:MeDIP Input 02 59
Track 3	BLEC:MeDIP 37
Track 4	BLEC:MeDIP 66
Track 5	BLEC:MeDIP 70

Input Data

- UC_Breast450k
- UC_Breast450k_MeDIP_Atlas

Step 21: Drag "UC_Breast450k" and "UC_Breast450k_MeDIP_Atlas" track lists

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh Data Filter: Select a filter...

- ROI Repository
 - Databases
 - ROI Repository - hg18
 - ROI Repository - hg19
 - All Annotations in Database
 - Tracks
 - Class: Affymetrix
 - Class: Agilent
 - Class: ENCODE
 - Class: ENCODE - T.f. Binding Sites
 - Class: Enhancer
 - Class: GC
 - Class: Gencode
 - Class: Gene
 - Class: Gene Model
 - Class: GeneModel
 - Class: Illumina
 - e200 Hs Methylation:450K
 - e200 Hs Methylation:Autosomes 450K
 - Hs Methylation:27K
 - Hs Methylation:450K
 - Hs Methylation:450K - No SNPs
 - UC_Breast450k:BranchSpecific450k
 - Class: Marker

Details

Attribute	Value
Group	ROI Repository
Database	ROI Repository - hg19
Name	UC_Breast450k:BranchSpecific4...
Description	
BigBed	none
BinWin	none

Input Data

UC_Breast450k
UC_Breast450k_MeDIP_Atlas
UC_Breast450k:BranchSpecific450k

Step 22: Drag the "UC_Breast450k:BranchSpecific450k" ROI track from "ROI Repository"

System/Network
Data
QC and Pre-processing
Genome
Transcriptome
Cistrome
Epigenome
Metagenome
Visualization
Help

Welcome to the Genboree Workbench! [Getting Started]

Data Selector
Refresh
Data Filter: Select a filter...

- Epigenome Informatics Workshop (May 2012)
 - Epigenome ToolSet Demo Input Data
 - Epigenomics Roadmap Repository
 - GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - Projects
 - GenboreeUser_project
 - Use_Case_01_GU
 - Use_Case_02_GU
 - Use_Case_05_GU
 - Use_Case_07_GU
 - Use_Case_09_GU
 - Use_Case_12_GU
 - Use_Case_13_GU
 - Use_Case_14_GU
 - Use_Case_18_GU
 - Use_Case_19_GU
 - Use_Case_20_GU
 - Use_Case_21_GU
 - GMT_Tutorial
 - JonathanMill_Lab
 - Public

Details

Attribute
View Link
Group
Name

Epigenome

- Random Forest
- QIIME
- QC
- Search for similar signals by correlation
- Analyze Signals
- Compute Similarity Matrix (heatmap)**
- Create Track Lists from Newick Tree
- Slice Epigenomic Data
- Analyze Signals in the Context of Epigenome Atlas

Input Data

- UC_Breast450k
- UC_Breast450k_MeDIP_Atlas
- UC_Breast450k:BranchSpecific450k

Output Targets

- GenboreeUser_database
- GenboreeUser_project

Step 23: Drag your database and project to "Output Targets"

Step 24: Select "Compute Similarity Matrix (heatmap)"

GENBOR
Baylor College of Medicine

System/Network Data QC and Pre-processing

Welcome to the Genboree Workbench

Data Selector

Refresh

- Epigenome Informatics Demo Output Data
- Epigenome Informatics Workshop (May 2012)
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Lists of Tracks
 - UC13_450K_Normal
 - UC13_450K_Tumor
 - UC13_Breast_450K_1
 - UC13_Breast_450K_2
 - UC13_Breast_MeDIP_Atlas
 - UC_Breast450k
 - UC_Breast450k_MeDIP_Atlas
 - UC_Breast450k_RRBS_Atlas
 - UC_Breast450k_WGBS_Atlas
 - UseCase12_Breast_RRBS
 - UseCase12_Breast_RRBS_Sun_20

Tool Settings

Output Database/Project:

Database/Projects Of Interest: GenboreeUser_database Group: GenboreeUser_group
GenboreeUser_project Group: GenboreeUser_group

Epigenomic Experiment Heatmap Tool

Analysis Name MeDIP_BrSpecific-Epigenor

Normalization Quantile

Aggregating Function Avg

Distance Function (dist) Euclidean

Hierarchical Clustering Function (hclust) Complete

Key ☒

Key Size 0.75

Height 14

Width 12

Trace None

Density Histogram

Dendrograms to display on heatmap Both

No Data Regions

No Data Value 0

Remove No Data Regions? ☐

Submit **Cancel**

Step 25: Name your job so you can identify it later

Step 26: Change the height to "14" and width to "12" of the image to be output

Step 27: In jobs involving MeDIP-seq experiments it is important to **not remove** regions that had no mapped reads

at Baylor College of Medicine.

BCM
Baylor College of Medicine

HGSC
HUMAN GENOME SEQUENCING CENTER

System/Network
Data
QC and Pre-processing
Genome
Transcriptome
Cistrome
Epigenome
Metagenome
Visualization
Help

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh
Data Filter: Select a filter...

- Epigenome Informatics Workshop (May 2012)
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository

Details

Attribute	
View Link	
Group	
Name	

Epigenome

- Random Forest
- QIIME
- QC
- Search for Similar Signals by Correlation
- Analyze Signals
- Compute Similarity Matrix (heatmap)**
- Create Track Lists from Newick Tree
- Slice Epigenomic Data
- Analyze Signals in the Context of Epigenome Atlas

Input Data

UC_Breast450k
UC_Breast450k_MeDIP_Atlas
UC_Breast450k:BranchSpecific450k

Output Targets

GenboreeUser_database
GenboreeUser_project

What have you done up to this step?

You provided the heatmap tool with two types of tracks: data tracks, and annotation tracks. The signal intensity from the “UC_Breast450k_MeDIP_Atlas” data tracks are projected over the “UC_Breast450k_BranchSpecific450k” annotation track (lineage specific CpG regions). The Euclidian distance metric is calculated for every possible pair of data tracks over the chosen regions, and the similarity matrix is presented visually as a heatmap.

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh

Data Filter: Select a filter...

- GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Lists of Files
 - Lists of Tracks
 - UC13_450K_Normal
 - UC13_450K_Tumor
 - UC13_Breast_450K_1
 - UC13_Breast_450K_2
 - UC13_Breast_MeDIP_Atlas
 - UC_Breast450k
 - UC_Breast450k_MeDIP_Atlas
 - UC_Breast450k_RRBS_Atlas
 - UC_Breast450k_WGBS_Atlas

Step 28: Repeat the heatmap analysis but now use the RRBS list of tracks

- UseCase13_450K_Tumor
- UseCase1_Brain_A

Details

Attribute	Value
Track 1	AK:RRBS 27 19
Track 2	AK:RRBS 27
Track 3	BAC:RRBS 12 73
Track 4	BAC:RRBS 49 76
Track 5	BAG:RRBS 49 71
Track 6	BAG:RRBS 49 68

Input Data

- UC_Breast450k
- UC_Breast450k_RRBS_Atlas
- UC_Breast450k:BranchSpecific450k

Output Targets

- GenboreeUser_database
- GenboreeUser_project

Welcome to the Genboree Workbench

Data Selector

Refresh

- Epigenome Informatics Demo Output Data
- Epigenome Informatics Workshop (May 2012)
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Lists of Tracks
 - UC13_450K_Normal
 - UC13_450K_Tumor
 - UC13_Breast_450K_1
 - UC13_Breast_450K_2
 - UC13_Breast_MeDIP_Atlas
 - UC_Breast450k
 - UC_Breast450k_MeDIP_Atlas
 - UC_Breast450k_RRBS_Atlas
 - UC_Breast450k_WGBS_Atlas
 - UseCase12_Breast_RRBS
 - UseCase12_Breast_RRBS_Sun_2011

Tool Settings

Epigenomic Experiment Heatmap Tool

Analysis Name RRBS_BrSpecific-Epigenom

Normalization Quantile

Aggregating Function Avg

Distance Function (dist) Euclidean

Hierarchical Clustering Function (hclust) Complete

Key ☒

Key Size 0.75

Height 14

Width 12

Trace None

Density Histogram

Dendrograms to display on heatmap Both

No Data Regions

No Data Value 0

Remove No Data Regions? ☒

Step 29: Name your job so you can identify it later

Step 30: Change the height and width of the resulting image

Note: In the case involving RRBS experiments you should leave the Remove No Data Regions box checked

Welcome to the Genboree Workbench! [\[Getting Started\]](#)

Data Selector

Refresh

Data Filter: Select a filter...

- GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Lists of Files
 - Lists of Tracks
 - UC13_450K_Normal
 - UC13_450K_Tumor
 - UC13_Breast_450K_1
 - UC13_Breast_450K_2
 - UC13_Breast_MeDIP_Atlas
 - UC_Breast450k
 - UC_Breast450k_MeDIP_Atlas
 - UC_Breast450k_RRBS_Atlas
 - UC_Breast450k_WGBS_Atlas

Step 31: Repeat the procedure but now use the WGBS list of tracks

- UseCase13_450K_Tumor
- UseCase1_Brain_A

Details

Attribute	Value
Track 1	AG:BS 03
Track 2	AL:BS 3 11
Track 3	Aorta:BS 03
Track 4	AT:BS 03
Track 5	BGM:BS 02 98
Track 6	BHM:BS 10 37

Input Data

- UC_Breast450k
- UC_Breast450k_WGBS_Atlas
- UC_Breast450k:BranchSpecific450k

Output Targets

- GenboreeUser_database
- GenboreeUser_project

Welcome to the Genboree Workbe

Data Selector

Refresh

- ▶ Epigenome Informatics Demo Output Data
- ▶ Epigenome Informatics Workshop (May 2012)
- ▶ Epigenome ToolSet Demo Input Data
- ▶ Epigenomics Roadmap Repository
- ▶ GenboreeUser_group
 - ▶ Databases
 - ▶ GenboreeUser_database
 - ▶ All Annotations in Database
 - ▶ Tracks
 - ▶ Lists & Selections
 - ▶ Lists of Tracks
 - UC13_450K_Normal
 - UC13_450K_Tumor
 - UC13_Breast_450K_1
 - UC13_Breast_450K_2
 - UC13_Breast_MeDIP_Atlas
 - UC_Breast450k
 - UC_Breast450k_MeDIP_Atlas
 - UC_Breast450k_RRBS_Atlas
 - UC_Breast450k_WGBS_Atlas
 - UseCase12_Breast_RRBS
 - UseCase12_Breast_RRBS_Sun_2011

Tool Settings

Epigenomic Experiment Heatmap Tool

Analysis Name WGBS_BrSpecific-Epigenomr

Normalization Quantile

Aggregating Function Avg

Distance Function (dist) Euclidean

Hierarchical Clustering Function (hclust) Complete

Key ☒

Key Size 0.75

Height 14

Width 12

Trace None

Density Histogram

Dendograms to display on heatmap Both

No Data Regions

No Data Value 0

Remove No Data Regions? ☒

Step 32: Name your job so you can identify it later

Step 33: Change the height and width of the resulting image

Note: In the case involving WGBS experiments you should leave the Remove No Data Regions box checked

Status of the jobs submitted can be obtained through Job Summary

The screenshot displays the Genboree Workbench interface. At the top, the 'GENBOREE' logo is on the left, and the 'BCM' (Baylor College of Medicine) logo is on the right. Below the logos is a navigation bar with tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. A left sidebar contains a menu with options: User Profile, Groups, Hosts, Jobs, and Request Feature. The 'Jobs' option is highlighted, and a red box around it has an arrow pointing to the 'Job Summary' option in the main content area. The 'Job Summary' option is also highlighted with a red box. Below this, a 'Tool Settings' window is open, showing the 'Job Summary' tool settings. The 'Settings' tab is active, displaying fields for Start Date (2013/7/18), End Date (YYYY/MM/DD), Sort Order (Newest first), and Group By (None). A red box around the 'Generate Report' button has an arrow pointing to a text box that says 'Select "Generate Report" to see Job Summary'. The 'Advanced Settings' section is also visible. The main content area on the right has sections for Details, Input Data, and Output Targets.

GENBOREE

BCM
Baylor College of Medicine

System/Network Data QC and Pre-processing Genome Transcriptome Cistrome Epigenome Metagenome Visualization Help

Use Genboree Workbench! [Getting Started]

User Profile Groups Hosts Jobs Request Feature

Atlas Tools Access BRL AUTO TEST FDACC

Job Summary

Tool Settings

Job Summary

Tool Overview

Settings

Start Date 2013/7/18

End Date YYYY/MM/DD

Sort Order Newest first

Group By None

Advanced Settings:

Generate Report Cancel

Select "Generate Report" to see Job Summary

Details

Attribute Value

Input Data

Output Targets

Genboree is built & maintained by the Bioinformatics Research Laboratory at Baylor College of Medicine.

Genboree is a hosted service. Code is available free for academic use.

BRL

HGSC
HUMAN GENOME SEQUENCING CENTER

You will get the following e-mail message when your job is completed

Hello

Your Compute Similarity Matrix (heatmap) job completed successfully.

Job Summary:
JobID - wbJob-epigenomicsHeatmap-e3CshO-5730
Analysis Name - EpigenomeExpHeatmap2013-10-09-00:42:01

Inputs:
1. Entitylist -
2. Trk -
3. Entitylist -

Outputs:
1. Db -
2. Prj -

Settings:
analysisName - EpigenomeExpHeatmap2013-10-09-00:42:01
dendograms - both
density - histogram
distfun - cor
hclustfun
height
key
keySize
naGroup
normalization
removeNoD
replaceNAV
spanAggFun
trace
width

- The Genboree Team

Result File Location in the Genboree Workbench:
<http://genboree.org/java-bin/project.jsp?projectName=Roadmap%20Epigenome%20Data%20Analysis>

Clicking on the link
will take you to the
project page
containing your
results

Click on the "Link to results" in
your Project page

Project News:

2013/10/16: Genboree User ran Epigenomic Heatmap Tool (MeDIP EpigenomeExpHeatmap2013-10-16-01 52 36) and the results are available at the link below.

- **Study Name:** MeDIP EpigenomeExpHeatmap2013-10-16-01 52 36
- **Link to results**

2013/10/16: Genboree User ran Epigenomic Heatmap Tool (RRBS EpigenomeExpHeatmap2013-10-16-01 51 41) and the results are available at the link below.

- **Study Name:** RRBS EpigenomeExpHeatmap2013-10-16-01 51 41
- **Link to results**

2013/10/16: Genboree User ran Epigenomic Heatmap Tool (WGBS EpigenomeExpHeatmap2013-10-16-01 50 04) and the results are available at the link below.

- **Study Name:** WGBS EpigenomeExpHeatmap2013-10-16-01 50 04
- **Link to results**

NOTE - the "Link to results" will appear in
project page only after all the jobs are
completed.

Table of Content: Epigenomic HeatMap

Study Name: MeDIP EpigenomeExpHeatmap2013-10-16-01 52 36

User: Genboree User

Date: 2013/10/16 02:04 CDT

Epigenomic HeatMap Plots

[Heatmap](#)
[Correlation plot](#)

Click on the heatmap to see comparison of breast 450K and Atlas MeDIP-seq

Newick Tree Visualizations

Equal Branch Lengths

Rows [\[PNG\]](#) [\[SVG\]](#)

Columns [\[PNG\]](#) [\[SVG\]](#)

Scaled Branch Lengths

Rows [\[PNG\]](#) [\[SVG\]](#)

Columns [\[PNG\]](#) [\[SVG\]](#)

Natural Log Scaled Branch Lengths

Rows [\[PNG\]](#) [\[SVG\]](#)

Columns [\[PNG\]](#) [\[SVG\]](#)

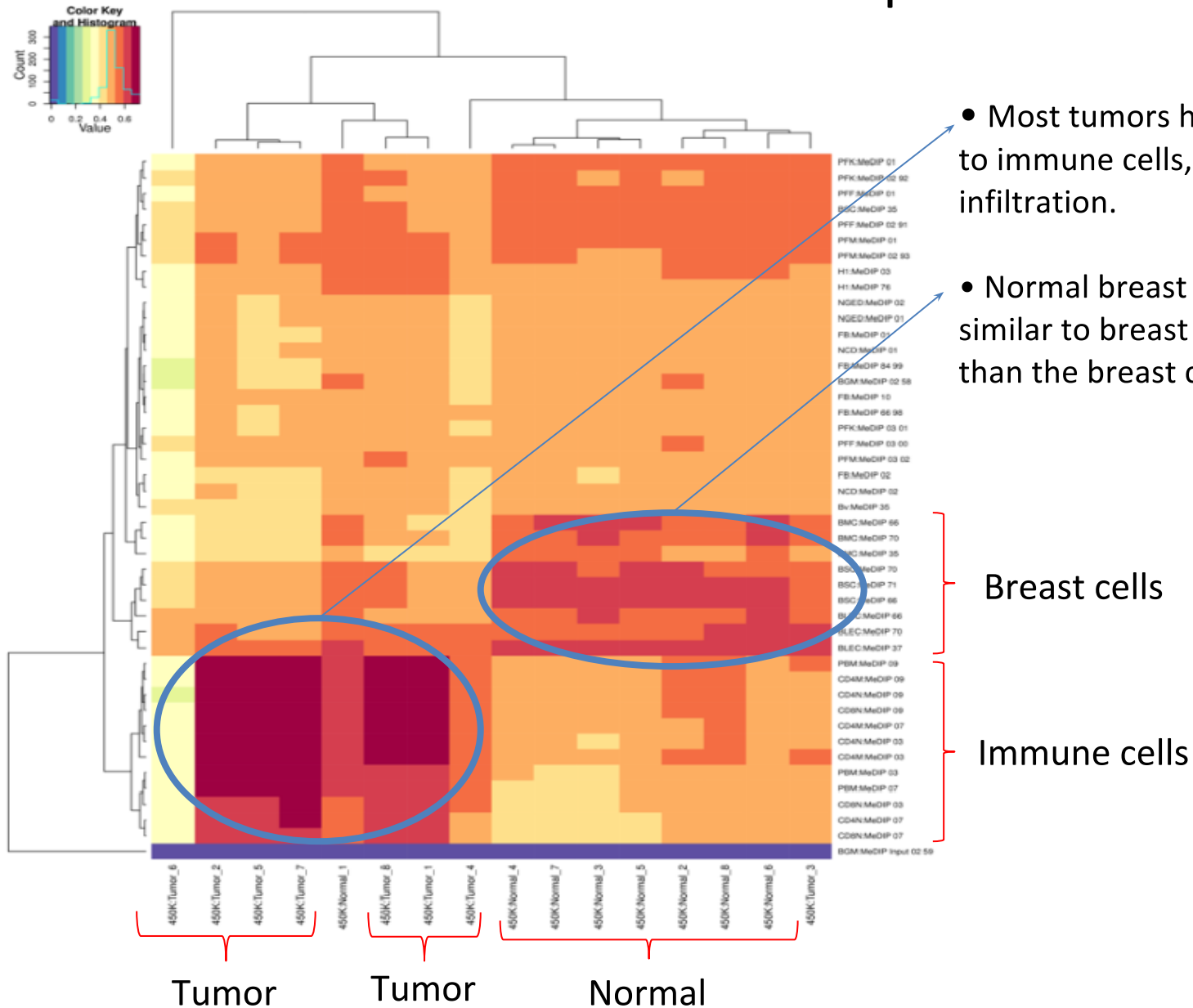
Log10 Scaled Branch Lengths

Rows [\[PNG\]](#) [\[SVG\]](#)

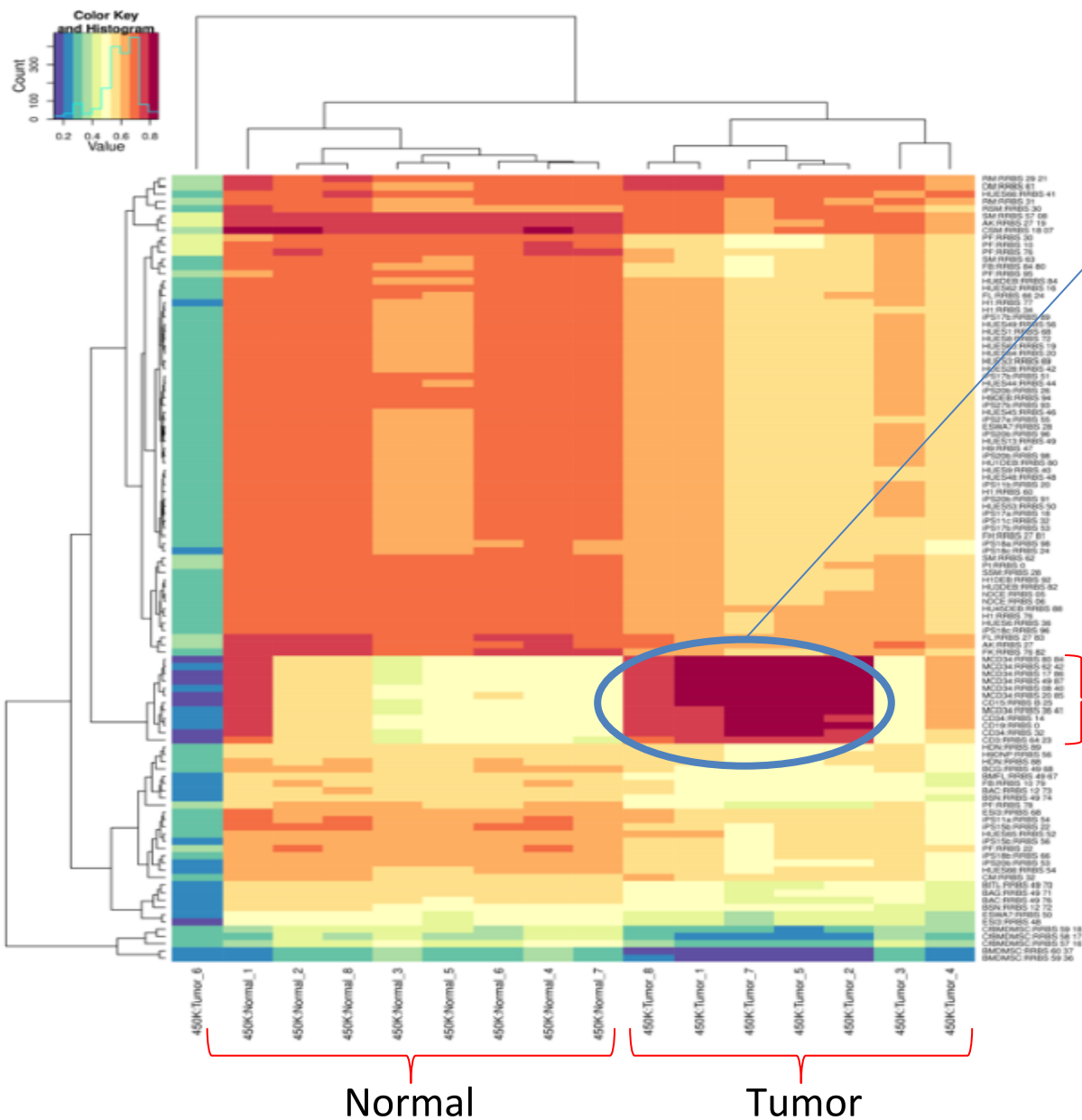
Columns [\[PNG\]](#) [\[SVG\]](#)

Note: Other types of plots are also generated by the heatmap tool. One of them is a correlation plot, which plots Pearson correlation metric between the tracks involved in this job. The other types of plots are hierarchical clustering generated dendrograms. The information in these plots is similar to what can be observed in the heatmap plot, but the representation may be more appropriate in some other cases.

Expected results: Comparison Breast450k vs Atlas MeDIP-seq



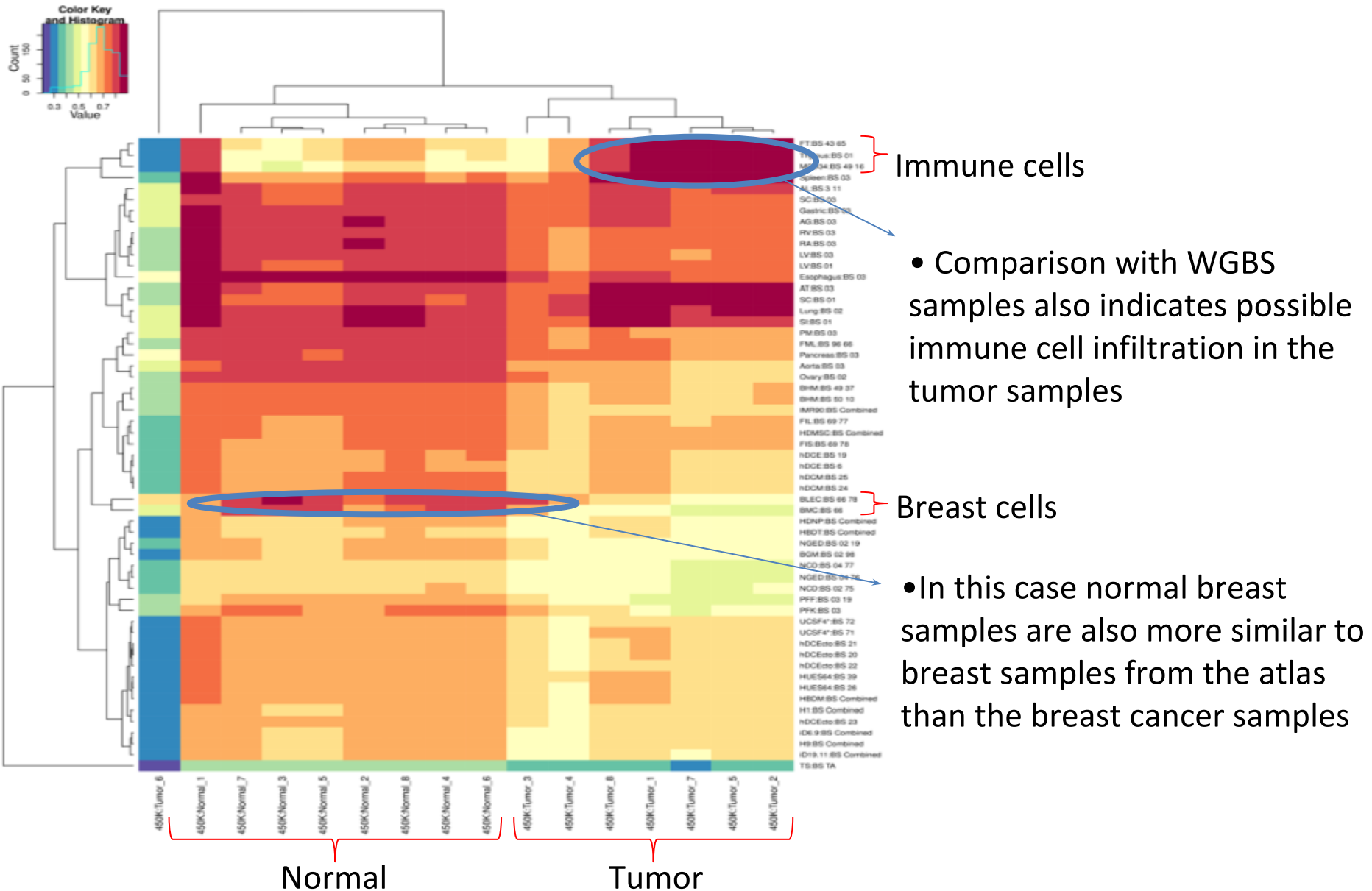
Expected results: Comparison Breast450k vs Atlas RRBS



- Possible immune cell infiltration in the tumors can also be observed in the comparison with RRBS atlas samples

Immune cells

Expected results: Comparison Breast450k vs Atlas WGBS



Help us improve Genboree. Please provide a comment or request feature.

The screenshot displays the Genboree Workbench interface. At the top, the 'GENBOREE' logo is on the left, and the 'BCM Baylor College of Medicine' logo is on the right. Below the logo, a navigation bar contains tabs for 'System/Network', 'Data', 'QC and Pre-processing', 'Genome', 'Transcriptome', 'Cistrome', 'Epigenome', 'Metagenome', 'Visualization', and 'Help'. The 'System/Network' tab is active, showing a sidebar with a tree view of system components. The 'Request Feature' option is highlighted with a red box. A red arrow points from this box to a 'Request Feature' dialog box that is open in the center. The dialog box has a title bar 'Request Feature' and a 'Settings' tab. It contains fields for 'User Name' (Genboree User), 'User Email' (andrewj@bcm.edu), and a 'Message' text area. 'Submit' and 'Cancel' buttons are at the bottom. In the background, the 'Details' table is visible, showing columns for 'Attribute' and 'Value', with rows for 'View Link' and 'Link to Project'.

GENBOREE

BCM
Baylor College of Medicine

System/Network Data QC and Pre-processing Genome Transcriptome Cistrome Epigenome Metagenome Visualization Help

Genboree Workbench! [Getting Started]

User Profile Groups Hosts Jobs Request Feature

Atlas Tools Access
BRL AUTO TEST
EDACC
Epigenome Informatics Dem
Epigenome Informatics Wor
Epigenome ToolSet Demo In
Epigenomics Roadmap Rep
GenboreeUser_group
GMT_Tutorial
JonathanMill_Lab
Public
ROI Repository
Targeted Atlases
genboree.bcgsc.ca
genboree.cbrc.jp
www.brain-research-lab.org

Data Filter: Select a filter...

Details

Attribute	Value
View Link	Link to Project
	GenboreeUser_group
	Use_Case_18_GU

Tool Settings

Request Feature

Tool Overview

Settings

User Name Genboree User

User Email andrewj@bcm.edu

Message

Submit Cancel

Genboree is built & maintained by the Bioinformatics Research Laboratory at Baylor College of Medicine.

Genboree is a hosted service. Code is available free for academic use.

BRL

HGSC
HUMAN GENOME SEQUENCING CENTER

Appendix: Methylation profiling techniques

- **WGBS (Whole Genome Bisulfite Sequencing):** Bisulfite is used to convert unmethylated cytosines into uracils in extracted genomic DNA. Bisulfite treated DNA is then sheared, amplified, and sequenced.
- **RRBS (Reduced Representation Bisulfite Sequencing):** Genomic DNA sample is digested with *MspI*, which cuts DNA at its recognition site (CCGG) independently of its methylation status. The fragments are then size separated and only those with size between 40bp and 220bp are selected, enriching for CpG rich regions. Bisulfite treatment followed by sequencing is then applied to selected fragments.
- **MeDIP-seq (Methylated DNA ImmunoPrecipitation followed by Sequencing):** Uses 5-methylcytosine specific antibodies to select for DNA fragments that contain methylated CpGs through immunoprecipitation. This process is then followed by high throughput sequencing.
- **Illumina Infinium 450k Array:** Genomic DNA is treated with bisulfite, fragmented, and amplified. DNA is then hybridized to a bead array. Each bead is covered with 50bp probes complementary to a region with one CpG. Some probes will match the CpG region assuming it was methylated, and had its cytosine converted to an uracil by bisulfite. Other probes will match the unmethylated version of the CpG. An extension reaction is then performed using fluorescently labeled nucleotides. Only probes with a perfect match to the interrogated locus will be extended. Fluorescence intensity is proportional to the level of methylation.